DOCUMENT RESUME

ED 329 555 TM 011 410

AUTHOR Terwel, J.
TITLE Implementation and Effects of a Program for Mixed Ability Teaching in Secondary Mathematics Education.
PUB DATE Apr 88
NOTE 21p.; Paper presented at the Annual Conference of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability Grouping; Academic Ability; Classroom Techniques; *Cooperative Learning; Foreign Countries; Grade 8; Junior High School Students; *Mathematics Instruction; Program Effectiveness; Program Evaluation; *Program Implementation; Secondary Education; *Secondary School Mathematics; *Small Group Instruction; *Teaching Methods
IDENTIFIERS Netherlands

ABSTRACT

A program designed to help teachers in the Netherlands put a form of mixed ability teaching into practice was evaluated for its implementation and effects. Implemented with 14-year-old students, the program consisted of whole class instruction by teachers, followed by guidance of small groups and individuals. Cooperative learning in small groups, mathematics set in real life situations, and different levels in the learning process were features of the program. The program was compared with a more traditional program for pre-test and post-test results. Antecedents were measured by a pretest and attitude scale. Processes were measured by a student questionnaire, the PERception of the Curriculum In Action (PERCIA)-scale. Outcomes were measured by a parallel version of the pre-test for learning results in the cognitive domain. Implementation of this program had a positive effect on learning results. Differences in implementation found in various classes had a significant effect on learning results. Teacher training and guidance can reduce variations in implementation, although class composition remains an important factor. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Implementation and effects of a program for mixed ability teaching in secondary mathematics education

ED329555

J. Terwel
University of Utrecht
The Netherlands

Abstract

This paper examines the implementation and the effects of a program for mixed ability teaching in secondary mathematics education. A comparison is made between this innovative program and a more traditional program. A pretest-posttest-control group design was used. Implementation of the program had a positive effect on learning results. Within the experimental condition, differences in implementation were found. These differences in fidelity of implementation significantly effect learning results.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the annual meeting of the AERA, New Orleans, 1988.

Mailing Adress:
J. Terwel
University of Utrecht
Faculty of Social Sciences
Department of Education (VOU)
Heidelberglaan 2
3584 CS UTRECHT
The Netherlands

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JAN TERWEL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

* This paper benefited from the helpful comments of Nathan Deen, Paul Herfs, and Karel Stokking. The research project was carried out by Rijkje Dekker, Paul Herfs, Wies Akkermans and Jan Terwel.

TM 011 410

Introduction

This paper presents the results of the last stage of a research project concerning the implementation and the effects of a mathematics program for 12-16 year-old pupils. (Terwel, Herfs, Dekker & Akkermans, 1987)

Implementation means "the process of putting a program into practice" (Fullan, 1982, Van den Akker, 1988, Stokking, 1988). The effects of this program are the affective and the cognitive results of learning. The program under study was designed in order to help teachers to put into practice a form of mixed ability teaching.

The implementation of programs for mixed ability teaching presents a number of problems as research has shown. The vast majority of teachers favours ability teaching (grouping) rather than mixed ability teaching in heterogeneous classes (Evertson et all, 1981, Kulik and Kulik, 1982, Marklund, 1984, Slavin, 1987). This was confirmed by a recent study carried out in the Netherlands. The majority of Dutch secondary teachers believe that mixed ability teaching in heterogeneous classes will hamper their task. They expect negative effects on the learning results of pupils who are at the extremes of the ability range. Most teachers use some kind of between-class ability grouping (Smeets and Buis, 1986).

In a large observation study in comprehensive schools in the United Kingdom, researchers found negative effects on the teaching and learning processes in mixed ability classes. They found many instances of low cognitive demands on pupils. Particularly the low and the high ability pupils often showed a lack of interest, boredom and off task behaviour. This result may be a consequence of the fact that teachers who teach heterogeneous classes do not always implement a form of mixed ability teaching or within-class differentation (Kerry, 1982).

The same situation exists in the Netherlands. Mooij (1987) found, contrary to his expectations, hardly any within-class differentiation in secondary education in the Netherlands. Teachers often claim to be implementing mixed ability teaching (within-class differentiation) but researchers rarely find evidence to support this. One possible reason for this discrepancy may be that schools may be granted additional facilities if they introduce innovations (including mixed ability teaching). Teachers will probably be tempted to suggest that more innovation has been achieved than is actually the case.

The various problems relating to the implementation of mixed ability teaching must be placed in the right context. At least three factors can be mentioned: (a) educational practice in large classes, (b) the lack of programs for mixed ability teaching and (c) the absence of adequate training and counselling of teachers.

The common practice of ability teaching also presents serious problems. Many 'mistakes' are made in the process of selection of pupils for the different streams or ability groups (Bos, 1974). Research has shown that the quality of the teaching and learning process in the lowest ability classes is usually very poor. The learning effects in these classes are small. The higher the proportion of low ability pupils in a class, the lower the quality of the cognitive transactions and learning results (Good and Marshall, 1984, Beckermann and Good 1981, Dar & Resh 1986). The

overall-conclusion to be drawn from the literature on between-class ability grouping is that this form of ability grouping has few, if any, benefits for pupil achievement (Slavin, 1987). However, some (meta-) analyses on ability grouping in secondary schools have claimed small positive achievement effects of between-class ability grouping, with high achievers gaining the most from the practice (Begle, 1975, Kulik & Kulik, 1982).

The majority of these findings underline the necessity of well designed programs for mixed ability teaching for those schools who have chosen for heterogeneous classes. Researchers may have an important task, namely, to describe and analyse the implementation and the effects of these programs.

The program: characteristics and theoretical background

The program for mixed ability teaching in this study was developed by the Dutch Institute for Curriculum Development (SLO). The evaluation was carried out by the University of Utrecht.

The program consists of materials for teachers and pupils (manuals for the teacher and textbooks/work materials for the pupils). It presents ideas, prescriptions, guidelines, examples and mathematical problems. This innovative program differs in many respects from more traditional programs in mathematics.

The central idea of the program is 'mathematics for all'. The curriculum developers tried to reconcile mathematics for the majority with more advanced mathematics for the minority (See also Freudenthal, 1980, Damerow and Westbury, 1983, Keitel, 1985, Terwel 1986 and Terwel in press, De Lange, 1985, Treffers, 1987). The discussion about 'mathematics for all' can be seen in the context of the more general, international debate about 'individual differences and the common curriculum' (Fenstermacher and Goodlad, 1983).

The implementation of the program creates a teaching-learning situation with the following main chararacteristics:

1. whole class instruction by the teacher;
2. guidance of small groups and individual pupils by the teacher;
3. cooperative learning in small groups;
4. mathematics set in real life situations;
5. different levels in the learning process.

The lesson (or series of lessons) generally consists of the following three successive parts:

A.Introduction:

The teacher introduces the problem to the whole class. He motivates the pupils by locating the mathematical problem in their world (real life situation). He presents the general outline of the problem, explores the various aspects of it and may suggest strategies for the problemsolving process.

B.Groupwork:

The pupils work in small groups. The teacher observes, guides, and encourages the problem solving process in the groups. He tries to solve

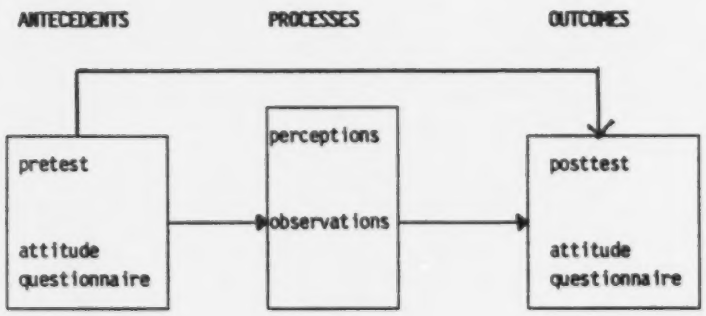
problems that originate in differences of speed or level between the pupils. When required, he helps pupils with individual learning problems.
C.Reflection and evaluation:
Following the groupwork, results and the actual process of the work in the small groups are discussed with the class. Special attention is given to the different strategies and solutions of the problem (the levels in the learning process).

The program is based on the theoretical ideas of Van Hiele (1982 and 1986) and Freudenthal (1973, 1973a, 1973b and 1980). On the basis of Van Hiele's level theory, Freudenthal proposes a new approach to individual differences. He points out that the levels in the learning process wich occur in classroom practice, can be used for the benefit of all. Differences between pupils are seen not as a hindrance but as an advantage. If pupils reflect on these differences (e.g. in levels, in solutions, in problem solving strategies) they can all reach a higher level of understanding. Mathematics exercised on a lower level becomes mathematics observed on the higher level. Pupils may apply rules subconsciously until they become aware of them. The results of these reflections may be the formulation of ideas and rules in more general terms. Cooperative learning in heterogeneous classes and in heterogeneous small groups is an important condition for this reflection process to take place.

Research model, research questions and hypotheses

The general research model for this study consists of three components.

Figure 1: Research model and variables



The research questions concern the implementation and the effects of the program. The four main questions are:

1. what is the affective and cognitive pre-knowledge of the pupils (antecedents)
2. is the program implemented in accordance with the intentions (transactions)
3. what are the effects of learning (outcomes)?
4. what are the relations between antecedents, transactions and outcomes?

The general hypothesis is: the more the characteristics of the program are realized the better the learning results will be.

From this general working hypothesis we derived two research hypotheses. The first hypothesis (A) refers to differences in implementation of the program within the experimental condition. The second hypothesis (B) concerns the comparison between the experimental group and the control group.

Hypothesis A.
In classes in which the teacher offers systematic whole class instruction and guidance and in which pupils cooperate in the problem solving process without engaging in off task behaviour, the learning results will be better than in classes in which those characteristics are not realized.

Hypothesis B.
In classes in which the innovative program has been implemented the learning results will be better than in classes in which a traditional program has been implemented.

This traditional program is with regard to mathematical content comparable to the innovative program. The main difference in content consists in the extent to which mathematical problems have been located in real life situations (contexts). Other differences relate to the relative absence of innovative characteristics described for the innovative situation. The experimental condition can be described as a form of cooperative learning and the control condition is a form of traditional whole class instruction.

Research setting and design

The innovative program was implemented in 22 classes in the second year of secondary education (fourteen-year-old pupils). Three schools and about 500 pupils were involved in this experimental condition. The control group, with a traditional mathematics program, consisted of about 300 pupils in eleven classes at two schools. Thus a total of five schools, eight hundred pupils in thirty-three classes were involved. All but one school were situated in the central part of the Netherlands. The experimental schools had more low-performing pupils and fewer high performing pupils than the control schools. Thus the experimental and the control groups

were not equivalent. To adjust for these differences gain scores, effects sizes and analyses of covariance were used.

At this stage of the research two different experimental designs were used. The design, which relates to hypothese A, was a correlational pretest-posttest design, with comparisons between classes within the experimental condition. For hypothesis B, a quasi-experimental pretest-posttest-control group design was used.

Methods and instruments

The research involved a variety of methods and instruments: tests, questionaires and observations. Below a short description of the instruments is given within the three components of the research model.

The antecedents were measured by a pretest and an attitude scale. The pretest was a performance test in mathematics. This test was used to assess the pre-knowledge of the pupils. The attitude scale was used to determine the attitude of the pupils to mathematics education, before the program was put into practice.

The processes were measured by a pupil questionnaire, the so called PERCIA-scale (PERception of the Curriculum In Action). This scale is a curriculum specific modification of existing measures of the perception of the learning environment (Fraser 1981, Fisher & Fraser 1983, Terwel 1986). In addition to this perception measure of the implementation of the program, observations were carried out.

The Outcomes were measured by a parallel version of the pretest for the learning results in the cognitive domain. The affective results were measured by an attitude scale (see antecedents).

In this paper the emphasis is on the data from the tests and the perception scale.

Processes: the implementation of the program

The fidelity of implementation was measured by the perception scale. PERCIA is designed to describe the implementation of the program as perceived by the pupils. We asked the pupils to rate the occurrence (or frequency) of activities (or situations) in the classroom. The following sample item gives an idea of design of the questionnaire.

	(hardly) ever				(nearly) always
The pupils help each other to find a solution					
This happens	1	2	3	4	5

The subscales on PERCIA were derived from the main characteristics of the program. Five clusters of PERCIA items were used. (The items relating to the different levels in the learning process do not form a specific cluster, but are part of cluster 1, 3 and 5)

1. Teacher whole-class Instruction and Climate (TIC)
The extent to which the teacher organizes introductions, reflections and discussions in the class as a whole. Some items about the social and affective climate in the class were included.
Sample item:
After the tasks are finished our teacher discusses them in class.

2. Guidance of small groups and individual pupils (GUID).
The extent to which the teacher gives guidance to small groups and individual pupils.
Sample item:
The pupils work together with help from the teacher.

3. Cooperation (COOP).
The extent to which the pupils help and understand each other.
Sample item:
We ask each other for help.

4. Task orientation (TASK).
The level of perseverance and commitment shown by the pupils without quarrels or distractions.
Sample item:
There is a great deal of talking and noise in the classroom.

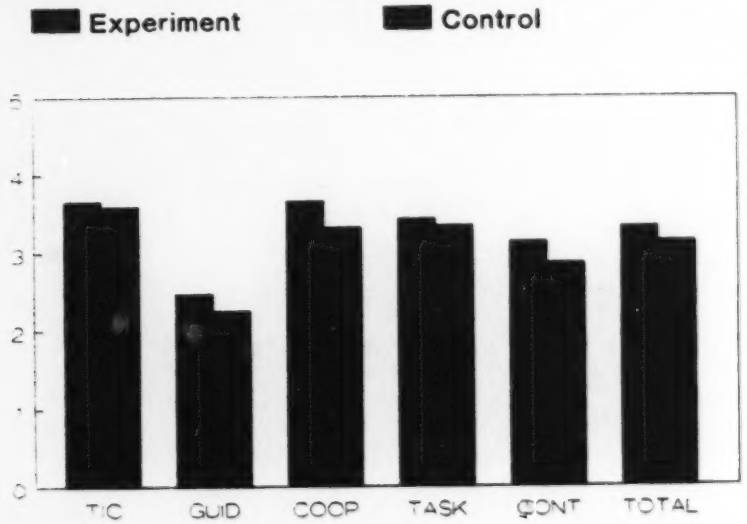
5. Mathematics in real life situations or contexts (CONT).
The extent to which mathematics is placed in real life situations.
Sample item:
The examples which are used in the lessons make mathematics attractive.

The reliability of the PERCIA-scale is reasonable to high. The alpha coefficients for the whole scale (54 items) was .85 and the Horst-coefficient was .94. The same holds true for the subscales (TIC through CONT). Each subscale had about 10 items. The alpha's for the subscales were about .70. The reliability of the mean class scores of all subscales was high: the Horst-coefficients range from .81 to 93. The PERCIA-results are presented in table 1 and a visual representation of these results is given in figure 2.

Table 1: Perceptions: means and analysis of variance between conditions

		TIC		GUID		COOP		TASK		CONT		TOT	
	N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
TOT.	835	3.64	.62	2.38	.50	3.54	.47	3.39	.55	3.04	.64	3.26	.36
EXP.	521	3.66	.59	2.47	.48	3.66	.45	3.43	.50	3.14	.64	3.33	.34
CONT.	314	3.59	.67	2.25	.52	3.32	.42	3.34	.62	2.86	.60	3.14	.34
F		2.7		40.0		127.1		5.8		38.8		56.26	
(df)		(1;833)		(1;833)		(1;833)		(1;833)		(1;833)		(1;833)	
sign		.10		.000*		.000*		.017*		.000*		.000*	
prop.var.		.003		.046		.13		.007		.043		.06	

Figure 2: Perception by conditon (exp/control) mean PERCIA-scores pupils



From table 1 and figure 2 we conclude that the innovative program has significant effects on the processes in classroom practice (as perceived by the pupils). In other words, the processes in the experimental condition are more in accordance with the intentions of the program designers than the processes in the control condition. However the difference between the conditions in terms of proportions "variance accounted for" is small. See the last row in table 1 which contains the ETA squared values. A closer look within the experimental condition shows interesting differences both at the school level and the class level.

Within the group of experimental schools we found significant differences in the extent to which the program had been implemented. School 1 shows a relatively low degree of implementation and school 2 has a high implementation score. The factor school accounts for 11 percent of the variance in the (total) perception score.

At the class-level (within the experimental condition) differences in implementation were found. The mean (total) perception scores differs from 2.88 - 3.63. These differences are significant and the proportion, variance accounted for the factor class is 34 percent (N = 22).

We devided the 22 classes into three categories according to their mean implementation scores (PERCIA-TOTAL). We found four classes with a low implementation score: below 3 points. Seven classes showed a high implementation score: above 3.4 points. The remaining eleven classes with scores from 3.0 to 3.4 had a medium implementation score.

Observations in all experimental classes also showed considerable differences between schools and between classes. The time spent on groupwork differed widely and so did the amount of off-task behaviour. School 1 had a high percentage of groupwork (72%) and a very high percentage of off-task behaviour (52%). School 2 and 3 spent about half of the time to groupwork and their off-task behaviour percentage was 5 and 15 percent respectively. In addition to the quantitative observations in all experimental classes we carried out qualitative observations in two classes (one class at school 2 and one class at school 3). These observations also showed differences in the implementation of the program.

In the class at school 2 the principle of "levels in the learning process" emerged more or less as intented. In the class at school 3 we hardly ever found this principle realized in practice.

From the perception scores and the observations we conclude that there were (large) differences in the extent to which the experimental programm was implemented. An important question is whether or not these differences relate to differences in learning results.

Antecedents and outcomes

The antecedents and outcomes (see research questions 1 and 3) were measured by a mathematics test. This test represented the content and the objectives of second year mathematics education in Functions, Relations, Graphs and Geometry. The mathematic problems in this test were placed in real life situations such as cycling, sports, camping, unemployment and bookcases. (for an example of a test-item see appendix 1) Pretest and posttest were parallel versions. The pretest had 16 items and the posttest had 16 parallel items plus 3 additional items to avoid ceiling effects (total 19 items).

The alpha coefficients were .74 and .83 for respectively the pretest and the posttest. The validity of the test was quarantee2d by careful content analysis of the program. The test items were authorized by the program developers. The tests represented the content of the program. In

the following the results of the pretest, posttest and gain scores are presented for different levels: class, school and condition. Only the pupils who took both tests were included in the analyses (N = 763).

Table 2: Mean pretest, posttest and gain scores by class (N = 33)

		Pretest		Posttest		Gain score	
class number	N	Mean	SD	Mean	SD	Mean	SD
11	19	28.6	9.0	35.5	10.7	6.9	8.8
12	10	22.0	9.7	32.6	11.5	10.6	5.7
13	15	18.9	5.0	31.2	10.8	12.3	10.3
14	19	22.7	6.2	26.4	7.9	3.7	6.3
15	18	25.0	8.1	33.8	11.0	7.8	6.3
16	16	22.6	8.9	28.8	10.0	6.2	8.2
17	11	23.9	10.2	33.9	14.1	10.0	6.5
18	13	18.2	7.1	13.9	7.8	-4.4	8.0
21	28	32.2	7.3	46.3	10.8	14.1	7.4
22	27	33.0	8.6	47.5	13.9	14.5	9.6
23	26	32.7	7.8	49.4	15.3	16.6	13.3
24	28	29.0	9.3	40.2	11.6	11.2	6.9
25	23	31.4	8.7	44.1	12.5	12.7	8.2
26	23	32.2	9.1	45.0	14.3	12.8	10.8
31	20	25.4	7.5	33.8	7.9	8.4	6.2
32	31	30.7	6.4	45.3	11.3	14.6	8.4
33	31	39.0	7.1	55.0	13.0	16.0	9.0
34	20	21.0	6.1	30.2	8.2	8.2	6.8
35	26	27.1	7.0	39.1	10.3	12.0	6.8
36	29	38.9	8.7	53.4	12.1	14.5	9.0
38	27	29.9	6.0	41.5	9.3	11.7	7.3
39	22	25.4	5.6	30.4	8.2	5.0	6.0
41	24	27.9	7.0	28.6	10.0	0.7	8.7
42	23	31.9	7.2	38.7	9.5	6.8	5.4
43	26	31.3	6.1	30.5	8.5	-.7	8.9
44	24	34.0	6.9	40.0	10.6	6.1	8.4
45	28	37.8	10.1	54.7	13.7	16.9	10.5
46	28	32.0	7.0	38.9	9.1	7.0	9.5
47	20	45.8	10.8	61.4	14.9	15.6	13.2
48	30	42.5	10.3	53.2	12.2	10.7	11.1
51	26	39.6	12.4	50.4	12.0	10.8	10.6
52	26	39.3	7.0	53.0	9.4	13.7	8.4
53	26	35.6	8.2	34.0	8.3	-1.6	7.1

From table 2 we conclude that there were large differences in pretest, posttest and gain scores between classes. Table 3 presents data at the school level.

Table 3: Mean pretest, posttest and gain scores by school (N = 5)

School	N	pretest		posttest		gain score	
		mean	SD	mean	SD	mean	SD
Total	763	31.7	10.1	41.6	14.7	9.9	10.1
School 1	121	23.2	8.5	29.7	11.9	6.5	8.8
School 2	155	31.7	8.4	45.4	13.2	13.7	9.6
School 3	206	30.6	9.0	42.5	13.7	11.8	8.3
School 4	203	35.4	9.9	43.2	15.5	7.9	11.2
School 5	78	38.2	9.5	45.8	13.1	7.7	11.0

The conclusion to be drawn from table 3 is that there were also differences in pretest, posttest and gain scores between schools. One of the experimental schools (school 1) had low scores on all three measures. Now we turn to differences between the conditions.

Table 4: Mean pretest, posttest and gain scores by condition (N = 2 resp. 3)

School	N	pretest		posttest		gain score	
		mean	SD	mean	SD	mean	SD
School 1-2-3	482	29.1	9.3	40.2	14.5	11.1	9.3
School 2-3	361	31.1	8.8	43.7	13.6	12.6	8.9
School 4-5	281	36.1	9.9	43.9	14.9	7.8	11.1

The data in table 4 show that there were differences in pretest, posttest and gain scores between the experimental condition (school 1,2,3) and the control group (school 4 and 5). This difference is significantly in favour of the experimental condition. The mean gain score of the experimental group is 11.1 point. In the control group the mean gain score is 7.8. Thus there is a difference of 3.3 points in the mean gain score between the conditions. If one expresses this difference in standard deviation units (with the SD of the posttest as a unit) one obtains an effect size of .22 (see Slavin, 1987 for the use of effect sizes in respect to gain scores). The difference between the mean gain scores (and so the effect size) becomes much larger if we exclude school 1 from the comparison (school 1 is a school with a high proportion of low achieving pupils and the teachers are relatively inexperienced). In that case the difference in mean

gain score is 4.8 and the effect size is .32. These effects of table 4 are represented by graphs in figure 3 and 4.

Figure 3: Learning results: pretest and posttest by condition

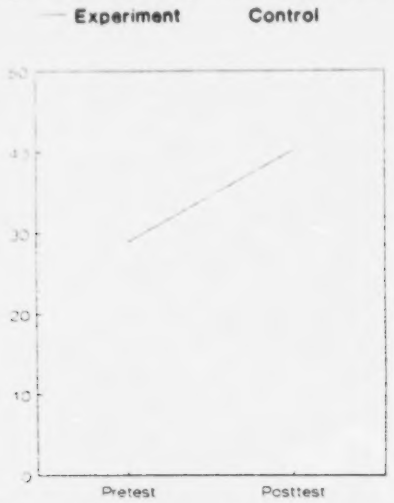
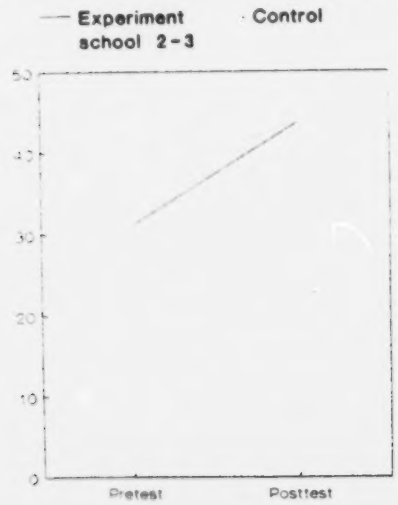


Figure 4: Learning results: pretest and posttest by condition



The main conclusion to be drawn from the comparison of the learning results in the two conditions is encouraging: the innovative program has a positive effect. Immediately below the relations between the degree of implementation of the program and the learning results will be analysed.

Relations between antecedents, processes and outcomes

In this paragraph the focus is on research question four: what are the relations between antecedents, processes and outcomes?

First we present the results on these relations within the experimental condition (22 classes and three schools). The correlations at the class-level are given in table 5.

Table 5: Correlations between perceptions and testscores within the experimental condition, class level (N = 22)

	TOT	TIC	GUID	COOP	TASK	CONT
pretest	.58*	.53*	-.17	.52*	.57*	.72*
posttest	.59*	.55*	-.20	.56*	.56*	.73*
gain score	.51*	.49*	-.14	.48*	.44*	.67*
proportion[1]	.50*	.49*	-.13	.32	.54*	.60*

[1] Proportion refers to the proportion medium and high performing pupils of a class according to their scores on the pretest.

The correlations between the class means within the experimental condition confirm our general hypothesis and the more specific hypothesis A. The better the characteristics of the program are realized, in the pupils' perception, the better the learning results are.

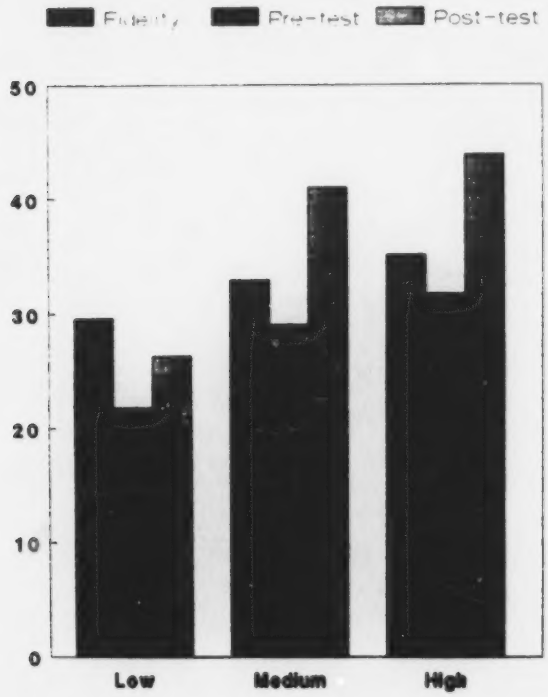
All correlations are reasonably high and significant, except the correlations in which guidance (GUID) is involved. This was not expected, although we had detected almost the same pattern at an earlier stage of our project (Terwel, 1986). It is possible that the pupils' need for guidance by the the teacher and the reaction of the teacher to requests for help are factors in this unexpected relation. Teachers may give more guidance in classes with a high proportion of low achieving pupils but the teachers' guidance cannot counterbalance the negative effects of a high proportion of low achievers.

The correlations in table 6 do not indicate the direction of the relations. Causal inferences cannot be made. The correlations between pretest and perceptions are of the same magnitude as the correlations between posttest and perceptions. In order to control for differences in pretest, a regression analysis was carried out. In this analysis the mean PERCIA (TOTAL) score was used. We assigned this mean perception score to every pupil in a particular class. This mean perception score of a class can be seen as the "treatment" for every pupil in that class. We carried out a regression analysis with the posttest as the dependent variable and the perception scores as the independent variable (while controlling for pretest). We found a significant but small increment in variance accounted for by the mean perception score over and above the variance accounted

and above the variance accounted for by the pretest. Thus we conclude that the degree of fidelity in the implementation has a significant effect.

The most striking differences emerge at the classroom level. As stated earlier we found wide differences in implementation between classes: low, medium and high fidelity. (See the paragraph "Processes: the implementation of the program".) Figure 5 gives a representation of these differences related to differences in pretest and posttest.

Figure 5: Implementation and effects: low, medium and high fidelity. Classlevel (N = 22)



From figure 5 we conclude that differences in implementation (fidelity) not only relate to pretest and posttest but also relate to gain scores (compare also tabel 5). From all these findings we may regard hypothesis A as confirmed.

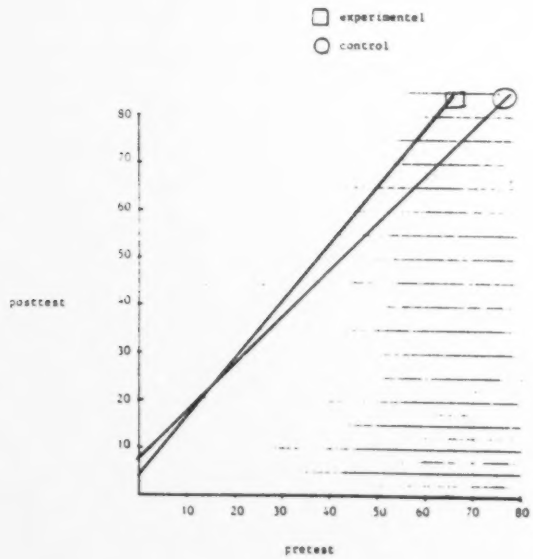
In addition to this conclusion we may say that school 2, the comprehensive school which realized mixed ability grouping most consistently, achieved the best results. School 3 which practiced a form of banding (limited ability range within a class) seemed at first to have produced slightly worse results. We compared school 2 with school 3.

However after controlling for initial differences (pretest) by analysis of covariance, the difference between the two schools was not significant.

We now turn to hypothesis B, which relates to the comparison between the experimental and the control condition. Regression analysis with the mean (TOTAL) perception score per condition as independent variable and posttest as dependent variable (while controlling for pretest) showed a significant effect in favour of the experimental condition. Thus hypothesis B was also confirmed. An analysis of covariance (with the pretest as covariate and condition as independent variable) also gives a significant result. We detected a small but significant interaction-effect between pretest and condition. We therefore turned a separate analysis of the results in the two conditons (Pedhazur, 1982, pp.469).

The two regression equations are:
Experimental condition: Posttest = 4.90 + 1.21 Pretest.
Control condition: Posttest = 7.80 + 1.00 Pretest. Figure 6 gives the two regression equations in a graph.

Figure 6: Regression lines: posttest on pretest by condition (N = 763)



By using the Johnson-Neyman technique the "areas of significance" can be determined. By these means we found that pupils with a pretest score of 23.6 or more (= 80 percent of the pupils) achieve better in the experimental condition than in the control condition. Below this point there is no significant difference between the two conditions. The conclusion is clear: for most of the pupils the innovative program had a positive effect on achievement. For the twenty percent who are low achievers there was no difference.

Discussion

Within the context of implementation the question of validity has to be discussed. The main question is: is the perception of pupils a valid measure of what really goes on in a classroom? Although perception scores may have an objective component, they may also contain a subjective, attitude component. We have not been able to isolate either component. Fortunately we have been able to do a form of "reality testing". We carried out systematic observations of "off task-behaviour" in all the experimental classes. The observation can be related to one of the implementation measures ,namely, task-orientation (TASK). The correlation between the two is -.54 at class level (N = 22). As expected, there is a negative correlation between pupils' perception of task-orientation (time-on-task) and time-off-task as observed by an outside observer. This correlation confirms the view that a mean perception score of a class does indeed give reliable and valid information about what is going on in that class. Although we do not have such cross-validations in respect of the other implementation dimensions, we would expect the same results there.

On the basis of alpha- and Horst-coefficients and this cross-validation we draw the conclusion that the implementation measures in this study are reliable and to some extent valid. The reliability and validity of the tests are also reasonable.

The main conclusion of this study is that the innovative program has a significant, but small, effect on learning outcomes.

We found differences in the extent to which the program was implemented within the experimental condition. This is an important conclusion from an implementation perspective. These differences have significant relations with learning outcomes. To overcome the problem of implementation teacher training and on-the-spot guidance of teachers is needed. But this is only one (important) side of the coin. Pupils are also a relevant factor. Implementation depends to some extent on class composition. The proportion of low and high performing pupils in a class seems to be an important factor (see also table 5). Further analysis and research are needed from a multi-level perspective. (Mooij, 1987).

Future directions

On the basis of the foregoing research we started a new research project involving mixed ability teaching in secondary schools: a project on adaptive instruction. This is the AGO 12-16 Project which focuses on the effects of a didactic arrangement of whole class instruction, combined with independent learning and cooperation between pupils. Special attention will be given to pupils at the extremes of the ability range. This means the focus will be more on within-class ability teaching (within-class differentiation) in the AGO-model, than was the case with the SLO-model (the Freudenthal-model) on which this paper was based.

References

Akker, J.J.H., van den (1988). Ontwerp en implementatie van natuuronderwijs. *Enschede: University of Twente, Department of Education.*

Beckerman, T. en Good, T. (1981). The classroom ratio of high- and low-aptitude students and its effect on a achievement American Educational Research Journal, 18, 317-327.

Begle, (1975), *Ability Grouping for Mathematics Instruction*: review of the empirical literature, California: Stanford University (SMESG Working Paper, no.17).

Bos, D.J., (1974 *Schoolkeuze-adviezen: resultatencontrole na vijf jaar*. Den Haag: Mouton.

Damerow, P. & I. Westbury (1985). Mathemathics for all: Problems and implications. *Journal of Curriculum Studies*, 17, 175-185.

Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal*, 23, 357-374.

Evertson, C.M., Sanford J.P. & Emmer, E.T. (1988). Effects of Class Heterogenity in Junior High School. *American Educational Research Journal*, 18, 219-232.

Fenstermacher, G.D. & Goodlad J.I. (1983). *Individual differences and common curriculum*. Chicago, University of Chigaco.

Fischer, D.L. & B.J. Fraser (1983). Validity and use of the Classroom Environment Scale. *Educational Evaluation and Policy Analysis*, 5, 261-271.

Fraser, B.J. (1981). Learning Environment & Curriculum Evaluation: a review. *Evaluation in Education: An International Review Series*. Oxford: Pergamon.

Freudenthal, H. (1973 a). De niveaus in het leerproces en de heterogene leergroep, met het oog op de middenschool. In: *Gesamtschule Conferentie 1973*, Amsterdam: APS/Muurses.

Freudenthal, H. (1973 b). *Mathematics as an educational task*. Dordrecht: Reidel.

Freudenthal, H. (1980). *Weeding and sowing*, Dordrecht/Boston: Reidel.

Fullan, M. (1982). *The meaning of educational change*. New York: Teachers College Press.

Good, T.L., en Marshall, S. (1984). Do Students Learn More in Heterogeneous Groups? In: P.L. Peterson, L.C. Wilkinson en M. Hallinan (eds.). *The Social Context of Instruction*. Orlando: Academic Press.

Hiele, P.M. van, (1982). Fasen en stadia in de ontwikkeling van het denken, zoals die door Piaget worden geconstrueerd, vergeleken met de denkniveaus geïntroduceerd door Van Hiele, *Pedagogisch Tijdschrift Forum voor Opvoedkunde*, 7, 207-218.

Hiele, P.M. van (1986). *Structure and Insight: A Theory of Mathematics Education*. Orlando: Academic Press.

Horst, P., (1949), A generalized expression for the realiability of measures. *Psychometrika*, 14., 21-31.

Keitel, C. (1985). Mathematik für alle: Pädagogisches Postulaat oder Gesellschaftlichte Anforderung. *Nieuwe Wiskrant*, 5, 48-59.

Kerry, T. (1982). The demands made on Pupils Thinking in mixed ability classes. In: M. Sands, & T. Kerry (Eds.). *Mixed Ability Teaching*. London.

Kulik, C.C. en Kulik, J.A. (1982). Effects of Ability Grouping on secondary School Students: a Meta-Analysis of Evaluation Findings. American *Educational Research Journal*, 19, 415-428.

Lange, J. de (1985). Mathematics for all is no mathematics at all. In: *Proceedings of the 5 th ICME Conferenz*, Adelaide, 1985.

Marklund, S. (1984). Mixed Ability Teaching versus Streaming. *Scandinavian Journal of Educational Research*, 28, 101-110.

Mooij, T., (1987), *International multi-level investigation into pupil behavior, achievement, competence and orientation in educational situations*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs (SVO).

Slavin, R.E. (1987). Ability Grouping and student achievement. *Review of Educational Research*, 57, 293-336.

Smeets, E.F.L., & Buis, Th. J.M.N. (1986). *Leraren over de eerste fase van het voortgezet onderwijs*. 's-Gravenhage: Staatsuitgeverij.

Stokking, K.M. (1988). *National Educational Policy and External Support systems as conditions for curriculum implementations*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans 1988.

Terwel, J. (1986). *Mathematics for All: between Dream and Reality*. Paper presented at the Annual Meeting of the AERA, San Francisco 1986.

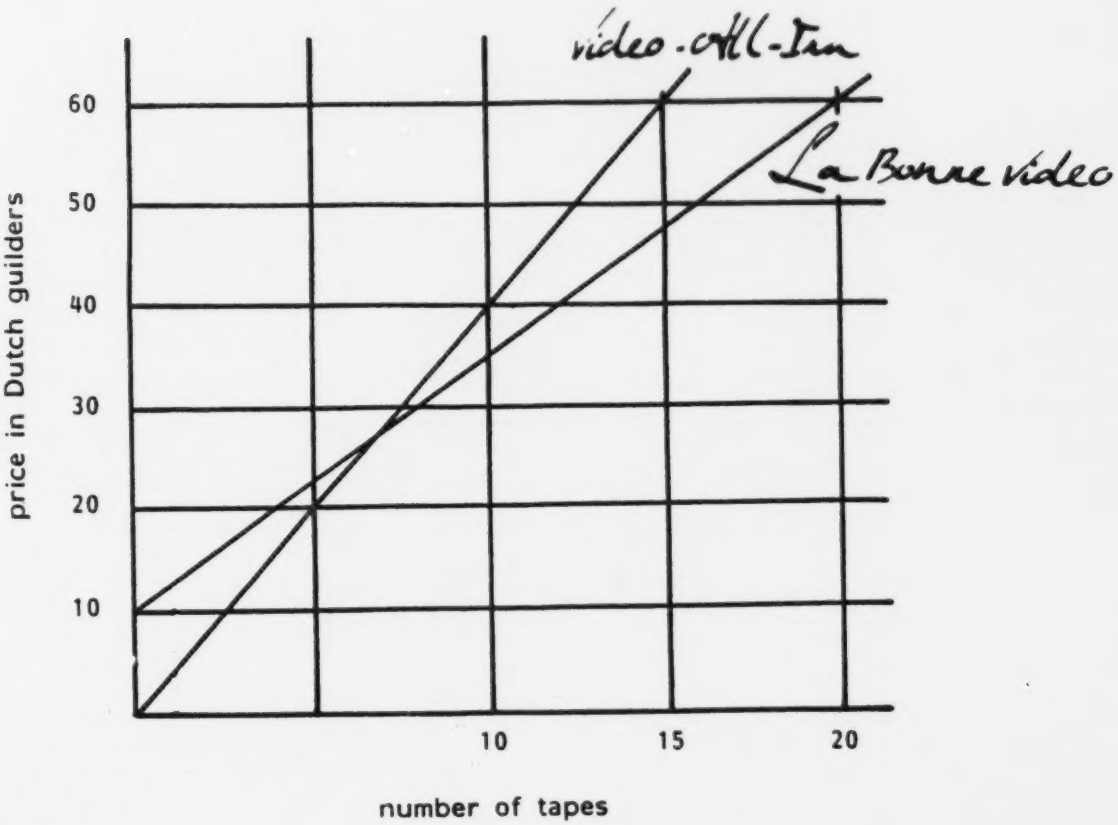
Terwel, J., Herfs, P., Dekker, R., & Akkermans, W. (1988). *Effecten van Interne Differentiatie*. Utrecht: University of Utrecht, Department of Educational Research.

Terwel, J. (In press). Real Maths in Cooperative Groups in Secondary Education. In: N. Davidson, *Learning Mathematics in Cooperative Groups*.

Treffers, A. (1987). *Three dimensions: A model of Goal and Theory Description in Mathematics Instruction - The Wiskobas Project.* Dordrecht: Reidel.

SAMPLE-ITEM

There are two shops in our district where you can rent video tapes: Video-All-Inn and La Bonne Video. In one of the two, you have to buy a membership card first. The other shop does not have these cards, you only have to pay per tape. Both have a daily charge (rate per day). The graphs showing the costs are given below.



Questions:

a) In which of the two shops do you have to buy a membership card?
b) How can you see that on the graph?
c) What does the membership card cost?
d) What is the charge per day per video tape in Video-All-Inn?
e) What is this charge in La Bonne Video (once you have a card)?
f) The lines in the graph cross at a certain point. What does this mean for people who often rent video tapes?
g) Video-All-Inn has a special offer this month: rent 3 tapes, pay for 2. What do you have to pay if you rent 15 tapes this month?
h) If you rent 40 tapes from La Bonne Video, how much do you pay?